CHAPTER 13

Universality and Diversity in the Vocalization of Emotions

Jörg Zinken, Monja Knoll, and Jaak Panksepp

Introduction

There is no doubt that prosodic channels play an important part in emotional communication (Bachorowski & Owren, 1995; Frick, 1985; Viscovich et al., 2003). A relationship between emotional experiences and acoustic properties of vocalization was already suggested by Darwin (1872), and it finds support from primate research (Jürgens, 2006).

The aim of this paper is to discuss universals in the vocal expression of emotion. In particular, we will share perspectives on the theoretical modeling of *universality* and *diversity* underlying research in this field. Our analysis will be based on diverse lines of evidence emerging from an increasingly substantial body of empirical research. Results from this research show an ambiguous picture, but they are most frequently interpreted as confirmation for the existence of universal mechanisms in the vocal communication of emotions. This research, briefly summarized in the first section, provides the background for discussions in the remainder of the chapter. In the second section, we scrutinize the notions of universality and diversity framing much of the current debate. In particular, we argue that misleading juxtapositions of universality versus diversity still seem to underlie the design and interpretation of empirical studies. Proposals for an alternative framing of universality and diversity are provided in the third section. In particular, we argue that universals of different types of affect vocalization need to be studied within their respective semiotic systems. The *symptomatic* communication of raw affect can be studied to discover evolved acoustic action patterns. Such patterns are continuous across species, and universal

across human communities. However, research into the cross-cultural recognizability of emotion from the voice has used speech stimuli obtained by asking speakers (mostly professional actors) to *pose* various emotions. Such *iconic* communication of affect can be studied as an aspect of advanced sign use in humans. Such sign use is diverse across cultures, but becomes universal in histories of engagement. In the fourth section, we discuss new questions that such a perspective brings into view, and we close with a set of conclusions in the fifth section.

Universality and Diversity in the Vocalization of Emotions: An Overview of Empirical Research

Results from Cross-Cultural Recognition Studies

The vocal expression of emotions has not received nearly as much attention as the facial expression of emotions. This is all the more true for analysis of universality and diversity in the vocal expression of emotions. Nevertheless, there is now a growing body of empirical research addressing these questions, including meta-analyses (Elfenbein & Ambady, 2002; Juslin & Laukka, 2003).[1]

Empirical research on the universality of vocal expression of emotions has largely focused on the ability of raters to recognize emotions expressed in speech samples produced by a member of a different culture (e.g., Albas, McCluskey, & Albas, 1976; Bezooijen, Otto, & Heenan, 1983; Scherer, Banse, & Wallbott, 2001).[2,3] Recognition studies consistently show an ambiguous picture. Virtually all studies report that emotions are recognized from voice with above-chance accuracy across cultures (Elfenbein & Ambady, 2002). At the same time, all studies show that the accuracy of recognition decreases with the cultural distance between the speaker and the rater (Elfenbein & Ambady, 2002; Scherer et al., 2001).

These results have received widely differing interpretations, depending on whether the authors were interested in universality or in diversity. For example, Scherer and his colleagues (Scherer, 1999a; 1999b; Scherer et al., 2001) tend to interpret the literature as providing support for the existence of a universal core of mechanisms for inferring emotions from the voice. Beside the fact that emotions are recognized with above-chance accuracy across cultures, they have found a similarity in confusion patterns. For example, the vocal expression of fear by a German-speaking actor was most frequently mistaken for sadness rather than any other emotion by both

[1]In most cases, cross-cultural studies of emotion recognition identify language with culture, so that the contrasted "cultural" groups are defined by the native languages. However, this is not always the case. In some studies, *culture* has been defined ethnically or geographically (Elfenbein & Ambady, 2002). The results have been similar to those where cross-linguistic diversity was studied. For example, Australian speakers of English were worse than American speakers of English at recognizing emotions from the voice of an American speaker.

[2]These studies aim to investigate people's ability to *recognize* emotion from voice. However, given the design of these studies, which require participants to choose an emotion from a list of alternatives, it seems more appropriate to interpret the results as indicating an ability to *discriminate* between emotions (Frick, 1985).

[3]The study of recognition of emotions bears parallels with the extensive literature on basic color terms using Munsell color chips: participants are asked to apply a local label to an array of foreign stimuli. It has been argued that this methodology produces rather than uncovers conceptual universals (Saunders, 1995).

German and Indonesian raters (Scherer et al., 2001).

However, the same results also indicate a systematic relation between cultural proximity and accuracy of emotion recognition. Notably, vocal expression of emotion by German-speaking actors in the study by Scherer and colleagues (2000) was recognized most accurately by speakers of German, less accurately by speakers of other Germanic languages (Dutch, English), less accurately still by speakers of other Indo-European languages (French, Italian), and least accurately by speakers of the only non-Indo-European language included, Bahasa Indonesia.[4] Although Indonesian raters still achieved above-chance accuracy, their accuracy rates were significantly lower than those of the German-speaking participants. Some evidence also suggests that the similarity in confusion patterns decreases with linguistic-cultural distance between speakers and raters. In a study on emotion recognition from the voice of English speakers by English, Spanish, and Japanese raters, the confusions in the judgments by Spanish speakers were similar to those of English raters, whereas those of Japanese raters were more dissimilar (Graham, Hamblin, & Feldstein, 2001).

The diversity evident from these studies becomes even more interesting when we take into account that the languages and cultures studied so far are not actually very diverse. For example, the most extensive cross-linguistic study in this area seems to be Scherer et al. (2001), who asked speakers of seven languages from nine countries to rate the emotions expressed in voice samples produced by German-speaking actors. However, of the seven languages included, six were Indo-European languages. With approximately 6000 languages currently spoken around the globe, this is hardly a representation of actual linguistic diversity, and it is unclear what conclusions about universality can be drawn from such a narrow and relatively homogeneous sample of languages.

Furthermore, only a few specific emotions have been studied systematically, and these usually relate to the different extant proposals regarding "basic" emotions. If universal mechanisms exist that guide the recognition of emotion from voice, or the vocal expression of emotion, these should surely exist for basic emotions. Given this restriction to a few basic emotions, the ambiguity of the results from recognition studies seems quite surprising. We would not expect such a result if the vocal expression of these emotions was itself grounded in universal mechanisms.[5]

An interpretation de-emphasising universal mechanisms might suggest that what these results show is the degree of familiarity of raters with the speaker's

[4]The only exception from this genetic trend was French-speaking Swiss, who achieved better recognition rates than speakers of Germanic languages (Dutch, English). The authors suggest that this is because of the fact that most Swiss speakers of French have some knowledge of German. More specifically, we would argue that it is the communication histories between French-speaking and German-speaking Swiss that leads to this result; cf. section The Relation between Universality and Diversity: A Systems Model.

[5]Indeed, when humans are asked to express a single word, such as *mom* or any other single-syllable word in four basic emotional intonations (happy, sad, angry, and scared), the forced choice emotional detection rate approaches 100% in American students (greater than 90% in four separate studies, Panksepp, 1985–1991, unpublished data). However, such emotional sounds could not be distinguished from EEG recordings from the cortical surface processed with sophisticated Event Related Desynchronization signal-detection algorithms, probably because the recording sites were far from the source generators (Panksepp, 2000).

culture. This notion is also supported by a much earlier study that was concerned with the effects of familiarity on the enhancement of emotion recognition within the same culture (Hornstein, 1967). It was found that the prosodic emotional communication between compatible college roommates improved during the first 3 months of living together. In contrast, incompatible roommates' emotional communication did not improve. In the remainder of the article, we will argue that recognition studies utilizing posed emotions can address such "developing universals," but not evolved universal "mechanisms."

Results from Cross-Cultural Studies on Infant-Directed Speech

Another area that has sought to combine research into prosody and emotional responsivity is that of speech directed to infants (e.g., Burnham, Kitamura, & Vollmer-Conna, 2002; Fernald, 1989; Fernald & Simon, 1984; Kuhl, 1994). The prosody of infant-directed speech (IDS) has been extensively analyzed and is characterized by higher pitch, exaggerated pitch contours, hyperarticulation of vowels, shorter utterances, and longer pauses (Burnham et al., 2002; Fernald & Simon, 1984; Stern, Spieker, Barnett, & MacKain, 1983; Trainor, Austin, & Desjardins, 2000). Independent raters of such speech perceived high positive emotional affect (Burnham et al., 2002; Kitamura & Burnham, 2003). These prosodic modifications are generally assumed to be universal, and several cross-cultural studies have found support for this (e.g., Fernald al, 1989; Kuhl, 2000; Scherer et al., 2001).

As a large proportion of research into IDS deals with the analysis of natural interactions, specific emotions (e.g., happiness, sadness) have rarely been addressed; instead content-filtered speech samples are normally rated for generalized emotional affect (such as positive, negative). What makes IDS an extremely relevant case study for the discussion of the expression of vocal emotion is the prelinguistic stage of the infant. This vocal expression of emotion toward infants is only possible through the modification of prosodic channels, as the lexical content of IDS is presumably not understood at this developmental stage. A second reason why we have decided to include IDS in this discussion is the fundamental assumption of universality in this area.

This viewpoint of universality is often maintained regardless of the fact that differences are apparent across cultures. For instance, differences were found in the way mothers express vocal affect between Australian English and Thai (Kitamura, Thanavishuth, Burnham, & Luksaneeyanawin, 2002). Thai is a tonal language, in which pitch is used to convey lexical meaning. Kitamura et al. (2002) found that Thai mothers compared to Australian English mothers restricted their pitch excursions in order to not disrupt tonal information. The authors concluded that instead of using heightened pitch to convey emotional affect as found in the English speaking mothers, Thai mothers may use different vocal characteristics (final particles to express mood and status) and maybe increased affective linguistic content.

Similarly, pitch contours in Mandarin Chinese (another tonal language) IDS are less exaggerated than in American

English IDS and are more similar to adult-directed speech (ADS) in Mandarin Chinese (Papousek & Hwang, 1991). This study also compared both IDS and ADS to foreign-directed speech (FDS), and interestingly found that Mandarin Chinese FDS had more exaggerated pitch contours than both IDS and ADS. In contrast to this, British English mothers' FDS does not show the same exaggerated pitch contours, and was found to be more similar to ADS, while IDS exhibited exaggerated pitch contours (Knoll, Walsh, MacLeod, O'Neill, & Uther, 2007). This shows that exaggerated pitch contours in tonal languages might have a different, more linguistic function compared to the nontonal languages, and we can therefore not assume that these acoustic features per se are indices of certain emotional states across all cultures. Many more examples of these slight but important differences in prosodic modifications across cultures in IDS exist. Japanese IDS has lower, less exaggerated pitch values than American IDS, but this is obviated by a higher affective linguistic content (Toda, Fogel, & Kawai, 1990; Viscovich et al., 2003). Another language that also exhibits lower pitch values in IDS but higher pitch values in ADS is Quiche (Mayan language, Bernstein Ratner & Pye, 1984). The authors explained their findings in terms of the utilization of pitch as an indication of status rather than affect.

The important aspect here is first that IDS shows slight differences in prosodic modifications across cultures, and that universality of the vocal expression of emotions per se can therefore not be completely maintained. We also suggest that prosodic modifications, regardless of how they are perceived by independent raters, might have a variety of different functions. For instance, in the case of the British English FDS, raters perceived low-pass filtered speech samples of FDS as more negative than ADS despite similar pitch contours and pitch values. The only difference that was found between those two speech conditions was the aforementioned hyperarticulation of vowels, which should have a linguistic but not an emotional function (Knoll & Uther, 2004). The research into IDS shows that regardless of features held in common in IDS across cultures, there are also differences that cannot be discounted and which would require an alternative, less simplistic explanation than the common universality versus diversity debate can offer.

The Relation between Universality and Diversity: Linear Models

Similarities and differences are both evident in the recognition of vocal expressions of emotion across cultures. This has been interpreted as showing that the vocal expression of emotions itself has both cross-culturally universal and diverse facets. Similarly, the acoustic properties of IDS show both similarities and differences across cultures. It seems then that instead of asking whether emotional expression is universal or diverse we should ask questions about the relationship between universality and diversity in emotional expression. This statement seems to be uncontentious and is endorsed so commonly in the introduction sections of research articles (e.g., Scherer et al., 2001) that it might even seem like a "shallow platitude" (Ellsworth, 1994).

However, explicit metatheoretical discussions of what we imagine the relation between universality and diversity to be are comparatively sparse (Manstead & Fischer, 2002).

We will argue in this section that the relation between universality and diversity has been modeled, more implicitly than explicitly, as the two poles of a single continuum.[6] This model is evident in the conviction underlying much of the empirical literature that if we find similarities in behavior across raters, then this will diminish the existence or at least the relevance of diversity. It is evident, conversely, in the argument that if we can illustrate diversity across cultures, then this decreases the convincingness, or relevance, of theories assuming the existence of universals. These *if-then* arguments are only valid if we can think of the truth about the expression of emotions as a point moving between the two poles on a continuum, or between two scales, so that an increase in universality necessarily decreases diversity and vice versa. A continuum model is also evident in some interpretations of differences in the recognizability of different emotions. For example, van Bezooijen and colleagues found that some emotions, such as sadness, were fairly well recognized from the voice across cultures, whereas others, such as joy, were relatively poorly recognized (Bezooijen et al., 1983). They concluded that some emotions are expressed in culturally specific ways, whereas others are universally expressed in the same way; in other words, emotions can fall on different places along a universality-diversity continuum.

A continuum model is evident also in the way universality and diversity may come together in the *process* of the vocal expression of emotions. This process is sometimes imagined as a sequential motion along the continuum, as represented in Figure 13–1.

The basic opposition between a universal and a diverse pole is here supplemented with a vertical metaphor: the evaluation of diverse practices being a surface behavior that is initiated by universal underlying mechanisms. The vocal expression of emotion is imagined to be initiated by universal evolved mecha-

[6]The figures discussed in this section are not intended as reconstructions of anybody's explicit theoretical views. Rather, they are meant as reconstructions of the nontechnical models and assumptions that underlie explicit scientific thinking (Black, 1962).

Figure 13–1. The expression of emotions as a sequence from universal to local steps.

nisms. In a second step, this expression is adapted to local, potentially diverse, display rules. This sequential model is evident, for example, in Scherer's metaphors of *push* and *pull* effects. Physiological changes in the organism constitute the push effects that initiate a vocal expression of emotion. Pull effects subsequently work on this expression to "suppress," "amplify," or "modify" it according to local display rules (Scherer, 1999b).

This model is also evident in the historical development of a division of labor between academic disciplines in the study of the expression of emotions. Much of the psychological literature treats the interest in diversity of emotional expression as belonging to the realm of anthropology and ethnology, which study local practices of showing emotions. Psychologists, on the other hand, have identified their subject as the universal mechanisms of emotion expression, which are thought of as an evolved potential that is itself universally shared across cultures, and possibly across species.

Accordingly, the interest in universality and the interest in diversity are effectively conceptualized as two distinct enterprises concerned with the different ends of the continuum. Interest in universality goes with a focus on evolved mechanisms, and interest in diversity goes with a focus on local practice. These views can be brought together by considering universality and diversity on different levels of generality. For example, the intrinsic capacities for infant musicality can lead to a host of culture specific learning principles that help weave a child into its local social ecology (Trevarthen, 2000).

Nevertheless, many authors still frame the debate as concerning universality *versus* diversity (e.g., Elfenbein & Ambady, 2002; Scherer, 2000), even when they explicitly acknowledge the importance of accommodating *both* universality and diversity. It seems that the research traditions of being either interested primarily in universals or in diversity are so strong that they still frame the current debate.

The continuum model can account for the empirical findings in the sense that it is able to incorporate both universality and diversity. However, this does not make it an appropriate model, and we will now consider that there are some problems associated with it. The model suggests that we can uncover universally evolved mechanisms of vocal emotion expression by identifying areas of overlap in the behavior (mostly recognition behavior) of members of different cultures. Those emotions for which we can find satisfying degrees of overlap (whatever these are) are emotions for which universal mechanisms of vocal expression exist. We have stripped away the diversity and identified the universal core.

We believe that this research strategy is not justified as long as intentional emotion portrayals are used. The reason is that intentional vocal communication and vocal affect bursts belong to two different semiotic systems (Borod, 2000). Affective vocalizations shared across species, which can be described as evolved universal mechanisms, belong to a symptomatic system. Posed emotion portrayals, on the other hand, require advanced abilities of sign use and belong to an iconic system. Both systems have their universals, but the relation between overlap in behavior in one system, and the structure and functions of the other system, is not as straightforward as the continuum model suggests. We will elaborate on this contention in the next section.

The Relation between Universality and Diversity: A Systems Model

In this section, we want to spell out an alternative model of the relation between universal and diverse factors in the vocal expression of emotions, represented in Figure 13–2. Again, this figure is not intended as a technical model, but as a figurative handle for the following discussion. In particular, we do not suggest, as it might seem from the figure: (a) that affect bursts occur independently of context; (b) that evolved universals are independent of evolving universals.

The gist of this model is that it separates affect burst behavior (the affective system) from intentional communicative behavior (the communicative system). Evolved mechanisms implicated in vocal bursts of emotion are part of the ancient affective systems that humans share with other primates and possibly other mammals (Panksepp & Bernatzky, 2002). There is presently little doubt that there are intrinsic emotional-vocal control mechanisms in the mammalian brain, with a great deal of cross-species coherence (Hauser, 1996; Jürgens, 1998). Vocal communication of emotions builds on the acoustics of affect bursts, but universality of such communication cannot be attributed to, and should not be sought in, just evolved mechanisms. In what follows we present this argument in more detail, discussing universality first in the affective system, and then in the communicative system.

Affect Vocalizations as Evolved Action Systems: Figure 13–2(a)

Emotional vocalizations occurring in all mammals can be described as evolved action systems (Panksepp, 1982; 2005). It would seem that such vocalizations, indicative of affective states experienced by other mammals, should also be a natural part of the vocal repertoire of the human species. These action systems would be part of what Scherer calls push effects.

Affective vocalizations might not so much relate to cognitive aspects of emo-

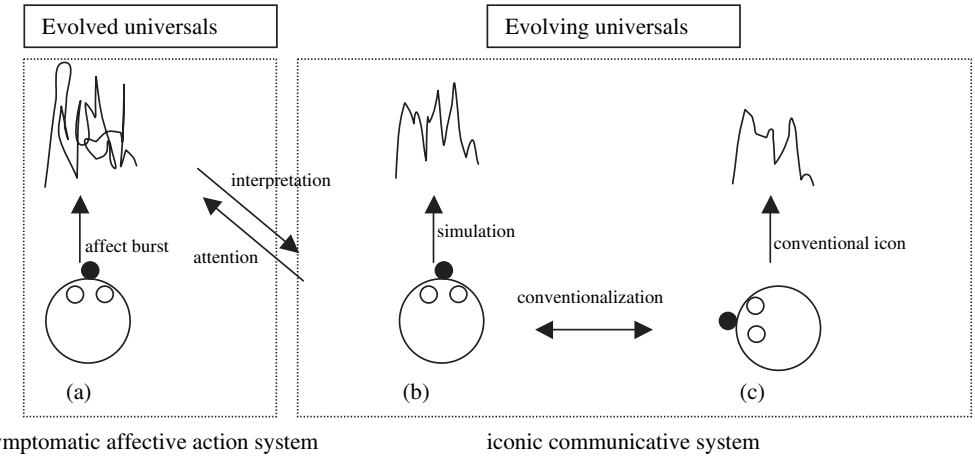

Figure 13–2. Systems model of the vocal expression of emotion.

tions such as anger or happiness, but rather to fundamental affective states that are part and parcel of motivational arousals shared by many species, states that we can refer to in English with words such as being enraged, joyous, cold, lustful, etc. (Panksepp, 2005). Of course, the terms of a particular language might be associated with various evaluations, expectations, or action "scripts" (Wierzbicka, 1991) that might be quite specific to that language. Nevertheless, we believe that many of the raw feelings that exist beyond words are shared across human cultures and across species. It is the symptomatic vocal expression of fundamental affective states, rather than the iconic vocal communication of emotions, that we would expect to exhibit similarities in acoustic properties across cultures (Juslin & Laukka, 2003; Panksepp & Bernatzky, 2002).

The recognizability of such raw emotional sounds is also reasonable from the point of view that certain emotional (or affective) states are associated with distinct physiological changes such as increased muscle tension and sympathetic arousal. For instance, tension in both the respiratory and laryngeal muscles can lead to louder and higher pitched speech (Frick, 1985). These physiological changes would presumably occur cross-culturally during the expression of certain affective states if these were accompanied by increased muscle tension and subsequent sympathetic arousal.

The existence of intrinsic brains circuits for a variety of basic emotions, which can yield vocal outputs (Jürgens, 1998), from essentially the same subcortical brain regions in all mammalian species that have been studied (Newman, 1988) provides powerful testimony for the existence of intrinsic emotional vocal control in all mammals. Perhaps the most compelling example is the squeal of pain. However, one critical question is whether these sounds are communicatory or merely by-products of emotional expressions.

There is increasing animal data that many of these sounds do contribute to communication and infection of emotional states among interacting animals. For instance, happy laughter type 50 kHz chirps in rats facilitate social and sexual choices (Burgdorf & Panksepp, 2006), and 22 kHz distress calls reduce the likelihood that colony members will leave burrows to explore and forage. The infantile separation-distress vocalizations in many species motivate maternal retrieval and attention, and hence have clear survival value (Panksepp, 2003). It is not hard to imagine that such emotional sounds, even after enculturation, will still have distinct kinds of emotional impact even in highly encephalized species where learned display rules tend to override instinctual ones. Indeed, it does remain possible that the roots of cultural practices such as musical expressions of emotions still rely on our capacities for making and appreciating intrinsic emotional sounds (Panksepp & Bernatzky, 2002) as our capacity to mold such raw materials into complex communications through the engines of culture.

Symptomatic vocalizations might be difficult to study in humans. One reason for this is the methodological difficulty of using words relating to complex categories such as *joy* or *sadness* in an attempt to study nonlinguistic affective states. Another reason is that human societies seem to impose constraints on the expression of affective states, so that humans, at least adults, have a tendency to cover up these states to make their

vocal expressions socially acceptable (cf. Scherer's notion of pull effects). The final usage or contribution of these acoustic cues to the communication of emotions would differ depending on cultural conventionalization processes, which could be very diverse, depending on how different cultures viewed the propriety and usefulness of emotional frankness.

Societal restrictions in the form of display rules are therefore one factor in the development of diversity. However, such social expectations are not the only, and probably not the fundamental, factor that leads to a diversity of vocal expression of emotions across cultures. The influence of society on human behavior is not a purely negative, constraining one. Rather, the motive for communication itself leads people to modify their voice, as we will argue in the following sections. Thus, as soon as intentionality becomes part of communication, diversity begins to impose new levels of complexity upon universality.

Attention to and Interpretation of Natural Signs (Symptoms): Figure 13–2(b) and (c)

Whereas the ancient affective action systems that produce affect bursts are very hard to study in the human species, this is not the case for the acoustic percept that an affect burst produces. The cognitive significance of vocal expression of emotion lies in the fact that the expression can be attended to and so itself becomes an input into cognitive processes (Jackendoff, 1996). When the acoustic percept produced in an affect burst is attended to, either by the producing organism itself or by another, we leave the system of affect burst behavior, and intrinsic communication, and enter the systems of intentional communicative behavior.

As a communicative process, the vocal expression of an emotional state is not the endpoint, but rather the starting point. It has been implicitly assumed in the research that others understand the expressed emotion in a natural way, maybe because they possess the same mechanisms for expressing emotion vocally, or because of the existence of universal "inference rules" (Scherer, 2003) that allow receivers to infer vocalizers' emotional states from vocal cues. However, we must keep in mind that the naturalness of an *expression* does not make its *interpretation* natural. Although a vocal expression with particular acoustic qualities might be a natural symptom of a particular affective state in all humans (and beyond), the interpretation of the expression *as* a symptom is not naturally given.[7] In order for a particular acoustic percept to be interpreted *as* an expression of happiness, the expression has to be attended to and understood. While this understanding might indeed be possible due to evolved inference systems in the case of the vocalization of raw affective states, human vocal communication of emotions clearly goes beyond that.

Imitation and the Conventionalization of Vocal Icons: Figure 13–2(b) and (c)

The vocalization of emotions in humans is not restricted to symptomatic expression. Most of the time, vocal expression

[7]Although fever has always been a natural expression (a symptom) of an inflammation in the organism, this *interpretation* of fever required the development of medical knowledge (Rodriguez, in press).

of emotions in humans, and possibly some other primates, also has an appeal function (Scherer, 1992). That is, humans use the voice *in order to* do something, in order to achieve a change in the state of others. Humans intentionally use their voices for communication. An understanding of the reasons for universality and diversity in vocal communication in humans therefore ultimately requires us to investigate the evolution of this communication system in its mutuality with (a) the intertwined evolution of voluntary control over vocalizations and of the phonatory apparatus (Ploog, 1992) on the one hand, and with (b) the evolution of understanding other minds on the other hand.

Crucially, in explaining possible universals as well as diversity in human vocal communication of emotions, we can no longer refer to evolved mechanisms. The reason for this is the flexibility inherent in the processes of imitation and conventionalization. This flexibility is due to the fact that any actual speech event is more complex than the properties that become semiotically interpreted by a listener (Bühler, 1934/1982). For example, the vocal expression of happiness involves communication through a variety of acoustic channels (e.g., Frick, 1985), such as pitch, rate, tension, etc. The interpretation of a vocal affect expression *as* expressing a happy state can become associated with one or several of these channels in the processes of simulation and conventionalization, coarsely represented in Figure 13–2(b) and (c). The intentional vocal communication of emotion, while initially a simulation of universal affect bursts, modifies the acoustic percept in a multitude of ways that cannot be reduced to social repressions. The iconic expression conventionally used in the vocal communication of an emotional state is constrained not only by societal rules, but also by system-internal considerations. For example, if the vocal simulations of happiness and nervousness are similar enough to be confused frequently, the conventional vocal expression of one or both of these emotions will change to make them more distinguishable.[8] This means that the motive for engaging with others through communication is itself a driving force behind the development of diversity. Such functional constraints are themselves universal, but they have the effect that overlap in the intentional vocal behavior of adult speakers across cultures will bear a complex relation to evolved affect burst systems. This positive motive for local developments of vocal communication is the reason why diversity cannot just be peeled away to reach a universal core. No known direct link leads from one to the other. Indeed, there may not exist one single path from one to the other, since propositional communication has a neural infrastructure of its own. Still there has to be some kind of interconnection between the two, perhaps in the fundamental motivational urge to communicate, which is heavily represented in limbic social-brain structures such as the anterior cingulate, which also regulate social motivation. However, it is worth considering that even lower brain functions, such as the seeking urge, driven by general purpose dopaminergic motivational systems, is a major force in promoting the desire to communicate.

[8]The intentional use of voice in communication does not mean that changes in the conventional vocal expression of emotion are intended. Intentional actions on the individual level lead to unintended changes in convention (Keller, 1994).

Discussion

Universality and Diversity Do Not Exclude Each Other

The major consequence that follows from the systems model we have suggested is that universality and diversity do not stand in opposition.

On the one hand, we are convinced that many affective states are, on a preconceptual level, shared by all humans. These shared affective states are surely more numerous than the modest six or so basic emotions often suggested as universal. This universality, however, does not at all diminish the fact that the expression of emotion, as soon as it involves elements of volition and creativity, might become very diverse across cultures. What is more, this diversity is not at all superficial, a package for a universal content. Rather, communication processes, including vocal communication, are a major force in the development of complex affective categories such as happiness or despair that constitute a culture's "forms of life" (Brakel, 1994). Human utterances are more than just expressions of a subjective state: they give this state a form, and thus enable its scrutiny and categorization (Humboldt, 1835/1994; Jackendoff, 1996). They become an input to cognitive and communication processes.

Conversely, diversity does not at all diminish the existence or relevance of universals. Imagine two cultures in which the conventionalization of the vocal expression of happy affective states has taken maximally diverse paths. Say, culture A has conventionalized increased pitch variation as a signal of such states, and has dramatized this acoustic property in vocal communication. Culture B has conventionalized fast speech rate as a signal of happy states, and has dramatized this acoustic property in vocal communication. Here we have maximal diversity: the two typical expressions of happiness do not overlap in terms of their acoustic properties. Nevertheless, if presented with vocal portrayals of happiness from a member of culture B, a member of culture A might still be able to correctly identify the intended state, because she will recognize it as a *possible* simulation of happy affect burst vocalizations. This will be even more so if she is faced with a choice between, say, happiness, sadness, anger, and fear. Diversity can be absolute—this does not diminish the universal potential provided by the acoustics of affect bursts.

In order to say anything about evolved universals of affect burst vocalization, those vocalizations have to be studied in all their complexity, including neurological substrates. Hence, studying affect burst vocalization might be difficult in humans, and simpler in other animals. For instance, the topic is already being addressed in primate research (Jürgens, 2006). The study of intentional vocal portrayal of emotions, on the other hand, can address a different kind of universals. This is so because intentional vocal communication is constrained by factors that are not themselves related to the affective system.

Developing Universals

The communicative model we have presented [Figure 13–2(b) and (c)] predicts that overlap in the vocal expression of communication develops through processes of conventionalization, that is, in

social contact. In this sense, universals of vocal communication of emotion are developing as networks of communication expand. These are universals that develop through histories of engagement. In this sense, diversity is ultimately not based in ethnic, geographical, or linguistic factors, but perhaps mostly in lack of mutual engagement. This is supported by research showing that cross-cultural recognition rates of vocal expression of emotions increase with the increase of telephone contact between countries (Elfenbein & Ambady, 2002). It is also consistent with the finding that within a country, minority groups are better at recognizing emotions expressed vocally by majority group members than vice versa (Elfenbein & Ambady, 2002).

Research into universality and diversity of vocal communication of emotion should take histories of engagement across groups into account.[9] For instance, the extent to which cross-cultural intercourse is resulting in imitation of emotional prosody should not be underestimated. People might consciously or unconsciously mimic the prosody of interlocutors, particularly where continuous engagement occurs, much in the manner of a "phonetic chameleon." However, this appears to be presently untested.

Units of Analysis

The basic reservation we have about much of the empirical research and its interpretation can probably be summarized thus. A communication system cannot be properly understood if the units of analysis are the individual, on the one hand, and the society, on the other. We also need to take the communicative process as the unit of analysis. Dividing individuals from society has, in the current research context, led researchers interested in universality to a fundamentally negative view of communication and sociality in general: what the public context of emotional expression does is to constrain the individual, place rules on her behavior, and force her to suppress or modify emotional expression. We have argued that this view conceals many interesting questions about universality and diversity in the vocal expression of emotions, for example, the development of universality in histories of engagement. This leads to two types of universality (Marková, 1987) in affective vocalizations: evolved universals in symptomatic systems, shared across species, and developing commonalities in iconic systems, shared within communities, which become universal as networks of communication expand.

Universals and Functionality

We might therefore reopen the question of what it means if a behavior or psychological state is universal or not. Why should vocal expressions of emotion be universal anyway? The general point that researchers interested in universality wish to claim is that vocal expressions

[9]The fact that it is engagement in communication rather than just exposure to communication is supported by a study showing that students of English as a second language with high proficiency in English were not better at recognizing emotions vocally expressed by a native English speaker than students with low English proficiency. The authors conclude that learning a foreign language in a classroom setting might not be sufficient to become sensitive to this channel of communication.

are universal because of phylogenetic continuity (e.g., Scherer, 1999b). This would mean that vocal expression of emotion must be, and must have been, an adaptive behavior, that is, it must be functional. Its *functionality* determines the properties of an emotional expression (Darwin, 1872). However, in the sequence model of vocal expression discussed in the second section, the local display rules also *render* vocal expression functional. A more positive view of vocal communication of emotion would suggest that the community, the culture of the speaker, is itself one of the environmental systems that influence ongoing social microevolution. In this perspective, then, there is mutuality, rather than a unidirectional influence, between the two systems discussed here: the emotional-affective action systems and the cognitive communicative systems.

Conclusions

Raw affective experiences are a gift of nature; they are not cognitively penetrable (Panksepp, 2005). However, the percepts that result from vocalization *are* penetrable for humans: they can be attended to, repeated, reflected upon, dramatized, etc. Diversity develops in communication systems because expressions of emotion involve multiple channels, including various cognitive ones, but intentional vocalization might utilize only a subsection of these. In situations of fear the voice might become tense, it might tremble, and so on. In attending to the vocalization, a listener might pick out one of the vocal qualities, and ultimately a tense voice can become associated with fear in one group, whereas in another group a trembling voice is the more typical expression of fear. Communication systems undergo processes of conventionalization and change that are in part independent from the universal constraints of the affective system. Communication is a way of engaging the world that is distinct in principle from such evolutionarily more basic systems (Panksepp, 2005, p. 162).

The human voice comes to convey many subtle changes in emotional state —at least to somebody who has a history of engagement with the speaker. It is misleading to narrow down the universal to only a handful of emotions, in the hope that they will be the major way in which emotionality is communicated in the voice. Rather, it seems to us that there is a lot that is universal in the communication of affective states, and at the same time there is a lot that is diverse: many affective states (tiredness, playfulness, exhilaration, fear, many shades of distress, and many shades of joy) are surely felt by all people (Panksepp, 2005). However, as soon as such states are communicated, not to mention given names and thus ordered into categories, it is increasingly less fruitful to look at these vocalizations for a universal core of emotionality. Emotions become semiotically formed in communication, and emotions are easily recognizable from the voice if hearer and speaker share a history of using the same communication system. For example, it is easy to recognize contempt from the voice of a person that one knows well. However, the recognition of contempt has been very low in recognition studies (Elfenbein & Ambady, 2002).

What follows is that universality in the vocal expression of emotions can be studied in two ways. Studies of biologically

entrenched action systems might be more successful with nonhuman animals and very early stages of communication development in our own species. Studies of the vocal communication of emotions in humans that have been or are becoming enculturated might more fruitfully focus on the development of universality in histories of social engagement.

Models trying to incorporate both universality and diversity have taken the form of unidirectional sequence models: An emotional state triggers a universal mechanism (a motor program for vocal expression, for example), which is then modulated according to local display rules. Fundamentally, these models use conduit-metaphors of communication (Reddy, 1979/1993), where the basic emotions are the substance that itself remains untouched, and just becomes wrapped into different cognitive-communicative packages by the local code that needs to be decoded by the listener. We have suggested an alternative model that accommodates the ambiguous results of empirical research and motivates new questions about the universality and diversity of vocal expressions of emotions.

Vocal expression of emotion is usually studied as a natural sign, with a self-evident meaning. However, the fact that we can study the acoustic properties of an auditory signal and establish acoustic regularities does not mean that the meaning of such a natural sign is self-evident. The fact that a certain percept (such as a particular pitch contour) is a natural symptom of a state does not mean that it is understood *as* a sign of this state. As long as it is not clear what the relation is between auditory perception and conscious states, the description of acoustic properties cannot be used as a consistent description of a meaning.

In sum, we have been arguing that universals of iconic vocal emotional communication develop in histories of engagement. If such a process of developing universals exists, it must necessarily build on yet other universal potentials of the brain and mind. One of these potentials is accessible in the acoustics of affect bursts. Others have to be found in consistent patterns of social cognition, which may become increasingly complex with increasingly complex manmade environments. Presumably, the universal tendencies that govern social cognition would have less variability in humans still living in natural environments, but they may vary substantially depending upon the ecological factors of those environments. Future research might attempt to address the development of universals in the intersubjective engagements that emerge in different sociocultural contexts and world environments, and the relation between such developing universals and evolved vocal action patterns.

Acknowledgments. We are grateful to Alan Costall for discussions on an earlier draft of this chapter. Jörg Zinken's research is supported by the European Union's sixth framework project "Stages in the Evolution and Development of Sign Use."

References

Albas, D. C., McCluskey, K. W., & Albas, C. A. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology*, 7(4), 481–489.

Bachorowski, J. A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic

properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219–223.

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language*, *11*(3), 515–522.

Bezooijen, R. v., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology*, *14*(4), 387–406.

Black, M. (1962). *Models and metaphors: Studies in language and philosophy*. New York: Cornell University Press.

Borod, J. C. (2000). *The neuropsychology of emotion*. New York: Oxford University Press.

Brakel, J. v. (1994). Emotions: A cross-cultural perspective on forms of life. In W. M. Wentworth & J. Ryan (Eds.), *Social perspectives on emotion* (Vol. 2, pp. 179–237). Greenwich, CT: JAI Press.

Bühler, K. (1982). *Sprachtheorie*. Stuttgart, Germany: UTB. (Original work published 1934)

Burgdorf, J., & Panksepp, J. (2006). The neurobiology of positive emotions. *Neuroscience and Biobehavioral Reviews*, *30*, 173–187.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*, 1435.

Darwin, C. (1872). *The expression of emotions in man and animal*. London: John Murray.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235.

Ellsworth, P. C. (1994). Sense, culture, and sensibility. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 23–50). Washington, DC: APA.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412–429.

Graham, C. R., Hamblin, A. W., & Feldstein, S. (2001). Recognition of emotion in English voices by speakers of Japanese, Spanish and English. *International Review of Applied Linguistics*, *39*, 19–37.

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Hornstein, M. (1967). Accuracy of emotional communication and interpersonal compatibility. *Journal of Personality*, *35*(1), 20–30.

Humboldt, W. v. (1994). Charakter der Sprachen. In J. Trabant (Ed.), *Wilhelm von Humboldt. Ueber die Sprache: Reden vor der Akademie*. Stuttgart, Germany: UTB. (Originally published in 1835)

Jackendoff, R. (1996). How language helps us think. *Pragmatics & Cognition*, *4*(1), 1–34.

Jürgens, U. (1998). Neuronal control of mammalian vocalization, with special reference to the squirrel monkey. *Naturwissenschaften*, *85*, 376–388.

Jürgens, U. (2006). *Common acoustic features in the vocal expression of emotions in monkeys and man.* Unpublished manuscript.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814.

Keller, R. (1994). *On language change. The invisible hand in language*. London: Routledge.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's

speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392.

Knoll, M. A., & Uther, M. (2004). "Motherese" and "Chin-ese": Evidence of acoustic changes in speech directed at infants and foreigners. *Journal of the Acoustical Society of America*, *116*(4), 2522.

Knoll, M. A., Walsh, S. A., MacLeod, N. O'Neill, M. & Uther, M (2007). Good performers know their audience! Identification and characterisation of pitch contours in infant- and foreigner-directed speech. In N. MacLeod (Ed.), *Automated taxon recognition in systematics:Theory, approaches and applications* (pp. 299-310). Boca Raton, Florida: CRC Press/The Systematics Association.

Kuhl, P. (2000). Language, mind and brain: Experience alters perception. In M. E. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 99–115). Cambridge, MA: MIT Press.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822.

Manstead, A. S. R., & Fischer, A. H. (2002). Beyond the universality-specificity dichotomy. *Cognition and Emotion*, *16*(1), 1–9.

Marková, I. (1987). The concepts of the universal in the Cartesian and Hegelian frameworks. In A. Costall & A. Still (Eds.), *Cognitive psychology in question* (pp. 213–233). Brighton, UK: The Harvester Press.

Newman, J. D. (Ed.). (1988). *The physiological control of mammalian vocalizations*. New York: Plenum.

Panksepp, J. (1982). Towards a general psychobiological theory of the emotions. *Behavioral and Brain Sciences*, *5*, 407–467.

Panksepp, J. (2000). The neurodynamics of emotions: An evolutionary-neurodevelopmental view. In M. D. Lewis & I. Granic (Eds.), *Emotion, self-organization, and development* (pp. 236–264). New York: Cambridge University Press.

Panksepp, J. (2003). Can anthropomorphic analyses of "separation cries" in other animals inform us about the emotional nature of social loss in humans? *Psychological Reviews*, *110*, 376–388.

Panksepp, J. (2005). On the embodied neural nature of core emotional affects. *Journal of Consciousness Studies*, *12*(8–10), 158–184.

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: The neuro-affective foundations of musical appreciation. *Behavioural Processes*, *60*, 133–155.

Papousek, H., & Hwang, S. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, *12*(4), 481–504.

Ploog, D. W. (1992). The evolution of vocal communication. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 6–30). Cambridge, UK: Cambridge University Press.

Reddy, M. J. (1993). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 164–201). Cambridge, UK: Cambridge University Press. (Originally published in 1979)

Rodriguez, C. (in press). Object use, communication and signs. The triadic basis of early cognitive development. In J. Valsiner & A. Rosa (Eds.), *The Cambridge handbook of socio-cultural psychology*. Cambridge, UK: Cambridge University Press.

Saunders, B. (1995). Disinterring basic color terms: A study in the mystique of cognitivism. *History of the Human Sciences*, *8*(4), 19–38.

Scherer, K. R. (1992). Vocal affect expression as symptom, symbol, and appeal. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental ap-*

proaches (pp. 43-60). Cambridge, UK: Cambridge University Press.

Scherer, K. R. (1999). Cross-cultural patterns. In D. Levinson, J. J. J. Ponzetti, & P. F. Jorgensen (Eds.), *Encyclopedia of human emotions* (Vol. 1, pp. 147–156). New York: Macmillan Reference.

Scherer, K. R. (1999). Universality of emotional expression. In D. Levinson, J. J. J. Ponzetti, & P. F. Jorgensen (Eds.), *Encyclopedia of human emotions* (Vol. 2, pp. 669–674). New York: Macmillan Reference.

Scherer, K. R. (2000). A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. *Geneva Studies in Emotion and Communication*, *14*(1), 1–4.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15.

Toda, S., Fogel, A., & Kawai, M. (1990). Maternal speech to three-month-old infants in the United States and Japan. *Journal of Child Language*, *17*(2), 279–294.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195.

Trevarthen, C. (2000). Intrinsic motives for companionship in understanding: Their origin, development, and significance for infant mental health. *Infant Mental Health Journal*, *22*, 95–131.

Viscovich, N. B., Jr., Pihan, H., Peery, S., Brickman, A. M., Tabert, M., & Schmidt, M. S., Jr. (2003). Acoustical analysis of posed prosodic expressions: Effects of emotion and sex. *Perceptual and Motor Skills*, *93*(1), 759–771.

Wierzbicka, A. (1991). *Cross-cultural pragmatics: The semantics of human interaction*. Berlin, Germany: Mouton.